Marchetti, B., Karsili, T., Chakraborty, P., & Matsika, S. (2017). Mechanistic insights into photoinduced damage of DNA and RNA nucleobases in the gas phase and in bulk solution. *Faraday Discussions*. https://doi.org/10.1039/C7FD00188F

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Mechanistic insights into photoinduced damage of DNA and RNA nucleobases in the gas phase and in bulk solution.†

Pratip Chakraborty,[a] Tolga N.V. Karsili,*[a] Barbara Marchetti,[b] and Spiridoula Matsika*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

DNA/RNA photohydrates represent a class of well-known biomolecular lesion formed by the absorption of near- to mid- UV light. They are formed via a photo-induced nucleophilic hydrolysis reaction in which water is split (via nucleobase sensitisation) into H + OH radicals. These nascent radicals can then add across C5=C6 - forming a saturation of the preexisting double bond. If un-repaired, such lesions can lead to mutagenic carcinogenesis - which is responsible for several forms of cancer. Using high-level electronic structure theory (CASPT2), we map the key excited state reaction paths associated with the reactivity of DNA (guanine and thymine) and RNA (uracil) nucleobases with water. At the outset, we consider the intrinsic reactivity in the isolated gas phase - in which the water (cluster) + chromophore complex is free from environmental perturbations. We then extrapolate the thymine nucleobase to the bulk DNA environment in aqueous solution in order to ascertain the relative importance for hydrate formation in a more complex biological environment. In this latter study we use high-level mixed quantum/classical (QM/MM: CASPT2/AMBER) methods.

1 Introduction

The photophysics and photochemistry of DNA/RNA nucleobases have been the focus of extensive experimental and theoretical studies in the past two decades.[1–6] Such nucleobases are π-conjugated systems which absorb light in the near-UV region. Upon UV-irradiation, the nascent excited electronic state can create harmful photoproducts that may result in photocarcinogenesis. Such processes are usually minimised by nature - ensuring nucleobase photostability. In the present context, the well-known photostability of canonical nucleobases is defined by their inherently short-lived excited states characterised by the low fluorescence quantum yields.[1,2,6] Such short excited-state lifetimes are manifestations of non-radiative relaxation *via* internal conversion (IC) to the ground electronic state. The excess vibrational energy is thereby dissipated as heat - which in the cellular environment is transferred to the bulk. Such ultrafast processes are now well-understood to involve conical intersections (CIs) between excited/excited and excited/ground states.[3,7–9]

In addition to non-radiative decay, CIs are now recognised as mediators of many of the other ultrafast processes observed in molecular photochemistry. Photophysical investigations have found a variety of mechanisms through which chemical evolution has ensured photostability. These include (but are not limited to) out-of-plane ring-deformations about C=C and C=N bonds, bond stretches along $\pi\sigma^*$ states and electron-driven H-atom transfer in nucleotides and base-pairs. In many of these cases, the photoexcited nuclear wavepacket on the excited state surface evolves towards a CI characterised by one of the aforementioned mechanisms, to revert back to its ground electronic state in an ultrafast manner. The excess energy is usually partitioned as heat which, in the bulk, is dissipated to the surrounding environment.[3,9,10]

Notwithstanding, prolonged exposure to harsh mid- to deep-UV irradiation can lead to deleterious photoproducts - which are commonly referred to as photo-lesions. The associated quantum yields of such photo-lesions is typically $< 1\%$,[11,12] but can nonetheless have profound consequences such as cancers if formed. The most common form of photo-lesion is the cyclobutane pyrimidine dimer (CPD). CPDs form *via* a photoinduced [2+2] cycloaddition between the C5=C6 bonds of adjacent pyrimidine bases. The subsequent adducts can alter the structure of bulk DNA/RNA and are known to interfere with the functions of polymerases. Fortunately such adducts are usually repaired by nucleotide excision repair but rare cases can cause mutations which lead to photo carcinogenesis.[5] In addition to UV-light, analogous lesions can be formed by the oxidative addition

[a] *Department of Chemistry, Temple University, Philadelphia, PA 19112, USA. Tel: +1 215 204 2102; E-mail: tolga.karsili@temple.edu, smatsika@temple.edu*
[b] *Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA.*
† Electronic Supplementary Information (ESI) available: Additional energies and potential energy profiles, along with cartesian coordinates for all geometries reported here. See DOI: 10.1039/b000000x/

of radicals to DNA/RNA. The well-known lesion, 8-oxoguanine (8-OG) is formed *via* the oxidation of guanine which contains a lower redox potential than its canonical counterparts enabling it to form base-pairs with adenine that can lead to harmful mutations. Recent experiments by Ravanat and co-workers have evidenced the photo-induced formation of 8-OG following near-UV excitation of guanosine in aqueous solution.[13] The authors speculate that photoexcitation of the guanine nucleobase leads to photoionisation, the nascent cation of which reacts with the surrounding water solution to form 8-OG.

Another class of lesions involves the formation of photohydrate adducts of the aforementioned DNA/RNA nucleobases in an aqueous environment.[14–18] UV-induced photohydrate adducts form when a water molecule adds across the C=C/C=N double bond of the pyrimidine/purine ring in a so-called nucleophilic hydrolysis reaction. Remsen et al. have shown that uridine photohydrates are formed following UV-irradiation of a single stranded R17-RNA.[16] Synonymous chemistry occurs in aqueous solutions of uracil and thymine. In both cases, the nascent hydrate adducts are stable at 37 °C in neutral aqueous solution.[17] In comparison, cytosine hydrate is highly unstable and reverts back to cytosine in neutral aqueous solution.[18] The formations of photohydrates have been speculated to be attributable to ultrafast IC and subsequent reactivity of the vibrationally excited ground state nucleobase, but this assumption has not been verified, and our work will show that it is not necessary.[19]

In the cellular environment, the formation of such hydrate adducts represents nucleic acid damage - which is facilitated by removal of the bases from DNA via glycosylic activity.[18] If unrepaired, the nascent lesions can lead to deleterious mutations and interfere with transcription. The thymine hydrate also partially inhibits the activity of DNA polymerase I which could lead to lethal mutations.[20] Though clearly an important lesion, a detailed mechanistic study of the formation of such harmful photohydrates is still lacking. In this work, we present high-level single- and multi- reference electronic structure calculations that detail a given nucleobase + water photoreaction in order to ascertain the extent to which such lesion processes compete with intrisic non-radiative decay of a given nucleobase. We start by outlining the reaction of guanine in a small model cluster of water molecules in an attempt at understanding the first-order effects for forming 8-oxoguanine via UV-irradiation. We then extend such water reactivity to uracil and thymine in order to characterise the experimentally observed uracil and thymine hydrates. Finally, we extend the thymine + water photoreaction to the bulk DNA environment in order to ascertain the extent to which UV-induced photohydrate lesions persist in the bulk.

2 Results and Discussion

2.1 The gas phase guanine + water photoreaction

This section will highlight the details of mechanism associated with the Guanine (henceforth G; structure shown in fig. 1(a)) + water photoreaction - to devise a plausible path for forming 8-oxoguanine (henceforth 8-OG - structure shown in fig. 1(c)). In so doing, a full quantum chemical description of solute-solvent interactions is required. To this end, a reduced solute-solvent model, comprising the G chromophore complexed to five proximal water molecule (fig. 1(a)), was constructed in order to reduce the computational expense whilst ensuring an adequate description of the dominant solute interactions with the proximal solvent water molecules. Our solute-solvent model cluster size and orientation was motivated by the elementary first step in forming 8-OG - which is widely accepted to proceed *via* a molecular rearrangement of the guanine photohydrate (8-hydroxy-7H-guanine, henceforth 8-HG). The optimised geometry of 8-HG is displayed in fig 1(b) which shows an OH moiety bound to the C8 position at an instinctive perpendicular angle with respect to the plane of the G molecule. It is therefore intuitive that at least one of the cluster water molecules be bound dispersively above the plane defined by G (G-1H_2O). Through previous studies on analogous systems,[21–24] five solvent water molecules are deemed sufficient to qualitatively describe all significant orbital interactions - ensuring that all solvent molecules are optimally aligned for an accurate description of the ensuing G + H_2O reaction whilst maintaining a reasonable computational expense. The forthcoming description is therefore steered towards the excited-state chemistry of G in a cluster of five proximal solvent water molecules (henceforth simply G-H_2O) - the ground state minimum energy geometry of which is displayed in fig.1(a). The assigned atomic numberings will be used henceforth when describing the relevant atom motions. Fig. 1 shows that the optimised geometry of G-H_2O comprises two terminal water molecules each of which contains hydrogen bonding to either N4 or N1-H. The returned HO17H—N4 and H_2O26—H-N1 hydrogen bond distances of, respectively, 2.2 Å and 2.5 Å are typical of such chromophores - reinforcing our cluster size choice. In the presently optimised configuration a water-wire extends from H_2O26 to H_2O17 and positions H_2O29 in an optimal C2 position so as to form a dispersive interaction with the π-system of G.

Table 1 Vertical excitation energies (with oscillator strengths in parenthesis) and state characters associated with the lowest three excited states of the G + water complex.

Transition	Excitation energy / eV	Character
S_1-S_0	5.00 (0.1406)	$\pi\pi^*$
S_2-S_0	5.52 (0.0030)	$n\pi^*$
S_3-S_0	5.65 (0.3232)	$\pi\pi^*$

Table 1 lists the Vertical Excitation Energies (VEEs) and oscillator strengths accompanying the first three excited states of the G-H_2O complex. The orbital promotions associated with the preparation of the various excited states are given in fig. 2 - which shows that the lowest three vertically excited singlet states involve orbitals localised exclusively on the G chromophore. States arising from such chromophore localised orbitals are commonly (and will henceforth be) referred to as locally excited (LE) states. The S_1 and S_3 states are both of $\pi\pi^*$ character, the participating orbitals of which contain substantially large spatial overlap - manifesting in a large oscillator strength. In contrast, the S_2 state is best characterised as an $^1n\pi^*$ state; its dominant orbital transition

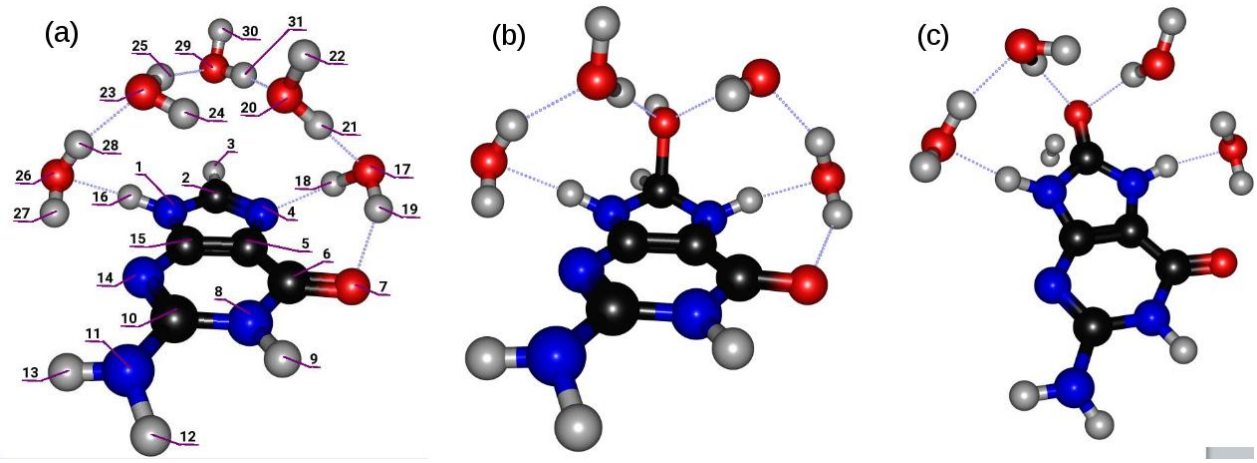


Fig. 1 Molecular structures associated with a) the parent guanine + water complex, b) the photohydrate adduct intermediate and c) the 8-oxoguanine product. The numbers in a) represent the assigned atomic numbering that we henceforth use when describing the atomic positions at which the reactivity is localised.

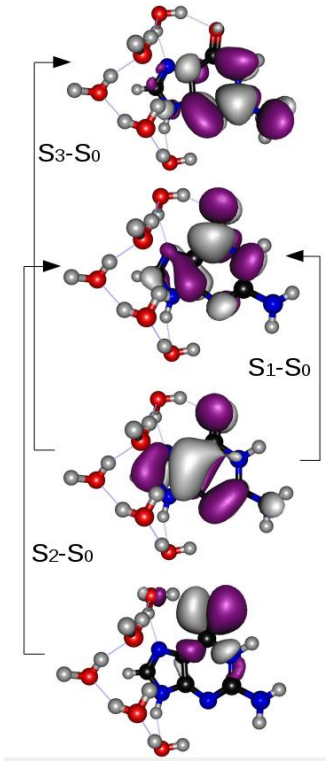
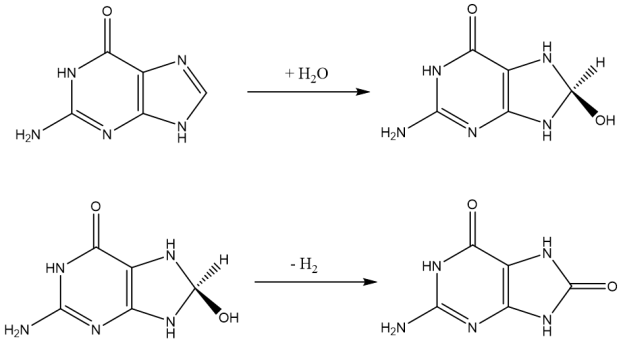


Fig. 2 Orbitals and orbital promotions associated with electronic transition to the lowest three excited states of the G + water complex.



Scheme 1

involves a $\pi^* \leftarrow$ n(O $2P_Y$) electron promotion. The characteristically weak oscillator strength of the S_2 (^{1}nπ^*) state is a manifestation of the poor spatial overlap between the participating n and π^* orbitals. The returned near-UV electronic configurations and energetic orderings listed in Table 1 are analogous to those of isolate G.[3]

The ensuing reactivity for forming 8-OG via the G + H_2O photoreaction is presented in a two-step sequential process (summarized in scheme 1) in which the initial step leads to the formation of the 8-HG intermediate followed by H_2 elimination from 8-HG to yield 8-OG. The potential energy (PE) profiles associated with each step are presented in, respectively, figs. 3(a) and 3(b). The first step is predicted to involve a two-step sub-path in which a solvent-to-chromophore electron-driven proton-transfer (EDPT - middle panel in fig 3(a)) initiates the photoreaction from the O17-H18 donor to the N4 acceptor. We note that as well as N4 there are other hydrogen-bond acceptors (e.g. N and CO) and donors (e.g. NH and NH_2) localised around the G chromophore. Our focus on EDPT at N4 is however motivated by the 8-OG end-product, which shows a hydrogenated N4 position. The inherent EDPT process is mediated by water-chromophore hydrogen-bonding and proceeds *via* an initial excitation to either the S_1 or S_2 LE states and is often referred to as proton-coupled electron transfer (PCET). Such LE states are then likely to couple to near-lying charge-transfer (CT) states - so characterised by electronic configurations in which an electron is promoted from a $2P_Y$ oxygen lone pair local to a proximal solvent water to a π^* orbital of the G chromophore. The S_1 relaxed PE profile along $R_{O17-H18}$ (middle panel in fig 3(a)) captures one such electronic configuration. A linear interpolation in internal coordinates (LIIC - left-

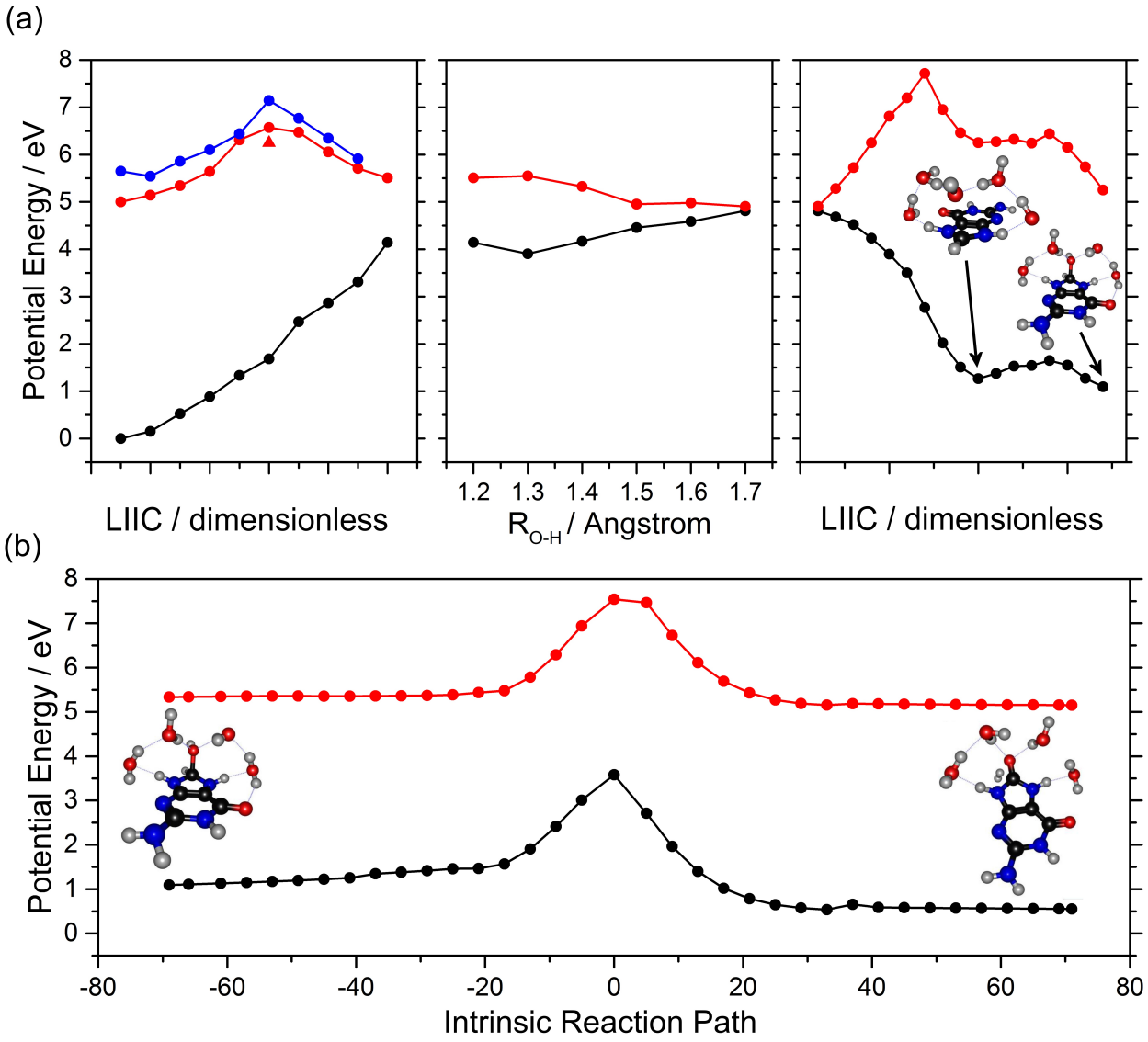


Fig. 3 CASPT2 potential energy profiles of the lowest three singlet states associated with the G + water photoreaction. The reader is directed to the main text for a fuller description of the details of the various profiles. The S_0, S_1 and S_2 states are, respectively, the black, red and blue curves.

most panel in fig 3(a)) shows the way in which the S_1(LE) configuration connects adiabatically to the S_1(CT) configuration. The returned LIIC shows a barrier height of 1.5 eV from the Franck-Condon (FC) geometry which necessarily represents an upper limit. The 'true' barrier height connecting the S_1(LE) to S_1(CT) was optimised (red triangle on the left-most panel of fig. 3(a)) and shown to reduce the barrier height by 0.3 eV. Such a relaxation in the barrier height, coupled with the recognition that the dominant nuclear motion is a H-atom migration, may lead to tunnelling beneath this barrier. Once populated, fig. 3(a) reveals that the S_1(CT) state is reactive with respect to O17-H18 bond elongation, representing proton-transfer from the O17-H18 hydrogen-bond donor to the N4 hydrogen-bond acceptor. The equivalent coordinate on S_0 is however unfavourable and shows a gradual increase in PE as a function of O17-H18 bond elongation. This favourable reaction on the S_1(CT) can be understood by recognising that the initial solvent-to-chromophore electron transfer creates a charge-separation between the donor water molecule and the acceptor G chromophore (i.e. creating G^- + H_2O^+). Migration of H^+ from H_2O^+ to G^- leads to a neutralisation of this charge separation - manifesting in a stabilisation of S_1(CT) as a function of $R_{O17-H18}$. In contrast, the closed-shell stable configuration of S_0 leads to an instability of the system upon O17-H18 bond elongation - manifesting in a rise in PE as a function of $R_{O17-H18}$. This subsequent rise and decline of PE in, respectively, the S_0 and S_1(CT) states inevitably leads to an energy crossing at $R_{O17-H18}$ ~1.7 Å. Such crossings between states of a common spin-symmetry comprise CIs and are therefore likely to mediate nonadiabatic transitions of the excited state population back to the S_0 PE surface via IC. If IC prevails at this long-range $R_{O17-H18}$ crossing, the system retains an excess energy of ~4 eV - which is nominally above many reaction barriers to various isomeric products on the ground state. We focus on two such reactions that may lead to the 8-OG - as outlined in reactions 1 and 2 in scheme

1. The first is the initial formation of 8-HG (reaction 1 in scheme 1) - which is predicted to contain a barrierless PE profile upon linear interpolation from the crossing geometry at $R_{O17-H18}$ ∼1.7 Å to the hydrate product - as displayed in the right hand panel of fig 3(a). The subsequent formation of 8-OG from 8-HG requires an elimination of H_2 as outlined in reaction 2 in scheme 1. Fig 3(b) shows the minimum-energy PE profile connecting 8-HG to 8-OG through an optimised transition state. It is noteworthy that this transition state is energetically below the crossing at $R_{O17-H18}$ ∼1.7 Å as well as in the near-UV excitation wavelength range.

Recent experiments by Ravanat and co-workers have evidenced the formation of 8-OG following near-UV excitation of guanosine in aqueous solution.[13] The authors speculate that photoexcitation of the guanine nucleobase leads to ionisation, the nascent cation of which reacts with the surrounding water solution to form 8-OG. In the present contribution, we show that it is indeed possible to form 8-OG *via* a one-photon (near-UV excitation) induced neutral photochemical path without the necessity for preparing a cationic form of G.

2.2 The gas phase uracil/thymine + water photoreaction

In this forthcoming section, we extend the above chromophore-water reactivity to that of uracil (U) and thymine (T) - in an attempt at understanding the mechanisms underlying the UV-induced formations of U- and T- hydrate products measured experimentally. Our initial structures comprise U and T chromophores complexed to a single water (H_2O) molecule - giving rise to dispersively bound uracil-water (U-H_2O) and thymine-water (T-H_2O) complexes. Our choice of a single water molecule is motivated by the need to describe the first-order water-chromophore reactivity without the bulk cluster effect. More specifically, the hydration reaction occurs via a nucleophilic hydrolysis reaction of a single water molecule, in which the nascent hydroxyl (-OH) and hydrogen (H) radical intermediates add across the C5=C6 double bond of the pyrimidine ring to form the 5-Hydroxy-6H-uracil (henceforth 5-HU), 6-Hydroxy-5H-uracil (henceforth 6-HU), 5-Hydroxy-6H-thymine (henceforth 5-HT) or 6-Hydroxy-5H-thymine (henceforth 6-HT) adducts as shown in fig. 4 (structures A, B, C and D, respectively). In order to ascertain the most stable hydrate adducts, the ground state geometries of structures A, B, C and D were optimised - the outcomes of which return 6-HU and 6-HT as lower energy adducts when compared to their C5-substituted analogues. This can be understood by considering the step-wise addition of H^+ and OH^- in which H^+ initially adds to either C5 or C6 via π electron-pair donation localised along C5=C6. Addition of H^+ to C6 delocalises the partial positive charge to C5 - rendering it electrophilic. The adjacent CO acceptor is a net π withdrawing group and thus destabilises the nascent C5-centred carbocation intermediate. In contrast, addition of H^+ at C5 renders C6 electrophilic - with a partial positive charge. The adjacent NH donor moiety serves to stabilise this electrophilic site via resonances - thus stabilising the C6-centred carbocation intermediate. The more stable C6-centred carbocation intermediate thus contains an enhanced rate and Boltzmann population. Charge recombination then leads to the OH^- to add to the C6-centre. This agrees with experimental results which show that 6-HU/6-HT are the observed products.[16,19,25] With these stabilising effects and energetics in mind, the forthcoming text therefore focuses on the U/T - H_2O reactivity in which the nascent hydrate products are 6-HU and 6-HT. Working back from these products, we then located a transition state (TS) on the electronic ground state - connecting 6-HU/6-HT to the corresponding dispersively bound chromophore-water reactant (see structure E in fig. 4). The eigenvectors associated with the single imaginary normal mode eigenvalue (see fig. S1 of the ESI) confirmed that the optimised TS involves the hydrolysis and subsequent addition of water across C5=C6 of U/T. The minimum energy TS was then used as the starting configuration for the construction of an IRC path - revealing the minimum energy reaction path connecting the possible reactants to the hydrate products. In so doing, the IRC path naturally reveals the possible lowest energy dispersively bound reactant structure from which we can tentatively start our studies on the excited-state chemistry associated with the U/T + H_2O photoreaction. Such initial reactant configurations are given in structures F and G in fig. 4. Whilst these do not represent the global minimum energy geometry returned from the IRC, they are the lowest energy dispersively bound structures that contain the optimal configurations for an ideal description of the dominant orbital interactions required for an excited state-nucleophilic hydrolysis reaction. The global minimum energy structures are however displayed alongside structures F and G in fig. 4 (F' and G'), are ∼0.2 eV more stable than F and G and represent a configuration in which the lone water molecules is hydrogen bonded to the N-H donor and CO acceptor sites. Though intuitively lower in energy, in solution such hydrogen-bonded configurations will co-exist with the dispersively bound configurations. The small energy difference between the minimum energy geometries and the orbital favoured geometries also serves to reinforce our choice for initiating our studies from structures F and G rather than from F' and G'.

Table 2 lists the VEEs to, and oscillator strengths associated with, the first two singlet electronically excited states of the U-H_2O and T-H_2O complexes. These agree well with previous studies on complexes of uracil with water.[26] VEEs of more states, in addition to more levels of theory, can be found in ESI. The orbital promotions associated with the preparation of the lowest two singlet excited states of U-H_2O are shown in fig. 5. The CASSCF (and subsequent CASPT2) energies reveal that the lowest two excited states absorb in the near-UV and are analogous to those of isolated U. The S_1 and S_2 states are of $^1n\pi^*$ and $^1\pi\pi^*$ character, respectively. The $^1n\pi^*$ state involves an orbital promotion from the $2P_Y$ carbonyl O-centred lone pair to a ring-centred π^* orbital and thus contains a characteristically weak oscillator strength on account of the poor spatial overlap of the participating orbitals. In contrast, the S_2 ($^1\pi\pi^*$) state contains a characteristically large oscillator strength manifesting from the appreciable spatial overlap between the participating ring-centred π and π^* orbitals. This strongly absorbing state is predicted to be the initially populated state leading to the photohydrate.

Armed with the knowledge of electronic state characters and associated energetics, we now describe the ensuing excited state

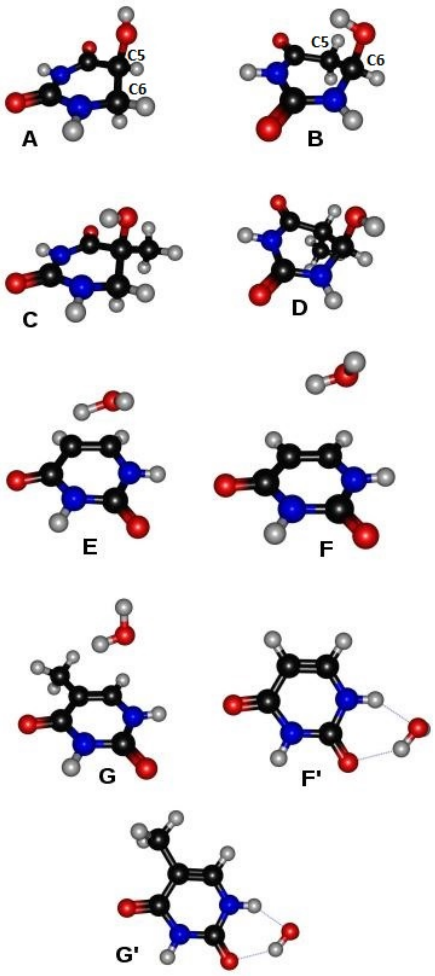


Fig. 4 Molecular structure associated with the various isomers and products of the uracil/thymine + water complex. The top two molecules depict the isolated uracil and thymine nucleobases; the assigned atomic numberings will be henceforth used in the main text.

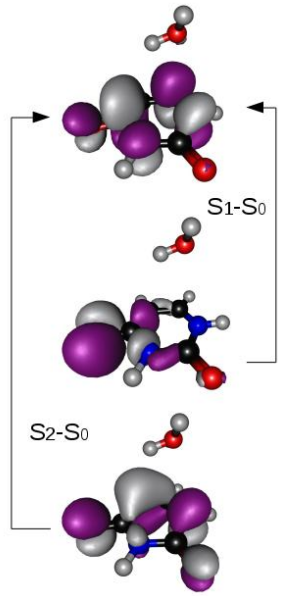


Fig. 5 Orbitals and orbital promotions associated with the lowest two excited electronic states of the uracil-water complex.

Table 2 Vertical excitation energies (with oscillator strengths in parenthesis) and state characters associated with the lowest two excited states of the T/U + Water complex at the CASPT2(20,14)/6-31G(d) level.

Transition	Excitation energy / eV	Character
Uracil		
S_1-S_0	4.96 (0.0007)	$n\pi^*$
S_2-S_0	5.20 (0.2309)	$\pi\pi^*$
Thymine		
S_1-S_0	5.00 (0.0007)	$n\pi^*$
S_2-S_0	5.12 (0.2241)	$\pi\pi^*$

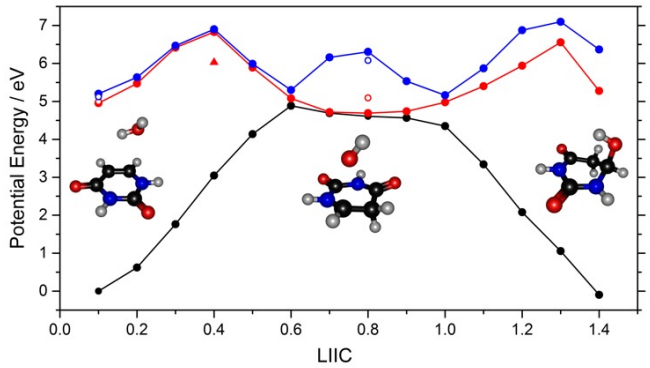


Fig. 6 CASPT2 PE profiles associated with the uracil/thymine + water photoreaction. The filled coloured circles represent the ab initio points associated with the uracil + water reaction whilst the open circles represent analogous points for various optimised points for the thymine + water reaction. The S_0, S_1 and S_2 states are, respectively, the black, red and blue curves.

reaction path for forming 6-HU from U-H_2O (henceforth $Q_{hydrate}$). Fig. 6 depicts the CASPT2 PE profiles associated with the lowest three singlet states, as a function of $Q_{hydrate}$. Equivalent PE profiles, calculated at the TDDFT levels of theory are displayed in fig. S2 of the ESI and are qualitatively similar to those obtained with CASPT2. For conciseness, we divide the description of the PE profiles along $Q_{hydrate}$ into two parts such that part 1) represents an initial LIIC path connecting the U-H_2O reactants to a minimum energy CI at $Q_{hydrate} \geq 0.8$ (henceforth LIIC1) and part 2) which represents a second LIIC path connecting CI to the 6-HU hydrate adduct (henceforth LIIC2). The dominant nuclear motions associated with LIIC1 involve the migration of the water centred H-atom to C5 whilst those of LIIC2 comprise the subsequent addition of the OH radical intermediate to C6. The profiles associated with LIIC1 reveal that at small displacements from the U-H_2O reactant (i.e. $Q_{hydrate}$ = 0 - 0.4), all states are bound and thus show no reactivity with respect to $Q_{hydrate}$. At intermediate to long-range geometries (corresponding to $0.4 \geq Q_{hydrate} \geq 0.8$) the S_1 and S_2 states become reactive with respect to $Q_{hydrate}$ - implied by the decrease in PE as a function of $Q_{hydrate}$. The initial rise and subsequent decline in PE leads to a local maximum on S_1 and S_2 at $Q_{hydrate}$ = 0.4 represents a change of electronic configuration from LE to CT character and is reminiscent of the aforementioned G + H_2O reaction. The S_1 and S_2 CT states are analogous in character such that both contain an electronic configuration in which an electron is promoted to a common ring-centred π orbital from either the O $2P_Y$ or O $2P_X$ orbital of water, respectively. Such an electronic configuration represents an inherent long-range charge-transfer which manifests in weak oscillator strengths for the S_1(CT) and S_2(CT) states. Notwithstanding, non-adiabatic coupling to the near-lying symmetry equivalent LE states can promote internal conversion from LE to CT. At near-threshold excitations to S_1, such couplings would need to surmount the aforementioned energy barrier at $Q_{hydrate}$ = 0.4 - which at first glance appears restrictive but merely represents an upper bound to the true energy at which such a transition would occur. Test optimisations of this local maximum were performed using TDDFT/B3LYP/6-31G(d) and by fixing the O-H distance of water at the value returned by the LIIC point at $Q_{hydrate}$ = 0.4 and allowing the remainder of the nuclear framework to relaxed to their respective minima on S_1. A LIIC path connecting the ground state minimum energy geometry to the returned relaxed geometry at $Q_{hydrate}$ = 0.4 (see fig. S3 of the ESI) reveals a 0.8 eV relaxation of this barrier (which now has a value of about 0.9 eV). Such an energy relaxation, coupled with the knowledge that LIIC1 predominantly represents PT motion, may indicate that this barrier may be surmounted by H-atom tunnelling. Any nascent population on S_1(CT) will then evolve towards the minimum energy CI and most likely undergo internal conversion to S_0 *via* nonadiabatic coupling at this CI. The profile associated with LIIC2 shows that states beyond (and including) S_1 show a rise in PE as a function of $Q_{hydrate}$. In contrast, the S_0 state continues to decline in PE in the range $0.8 > Q_{hydrate} > 1.4$ - the final geometry of which represents the 6-HU hydrate adduct (the actual 6-HU product is point $Q_{hydrate}$=1.5 but this could not be converged with CASPT2. see ESI for pathways including that point). Therefore, following internal conversion at the CI and in the limit of the 1-dimensional PE profile given in fig. 6, the evolving population may bifurcate and form nascent 6-HU products or reform the ground state molecule - but we recognise that in reality an excess of 5 eV following IC at this CI may surmount reaction barrier associated with orthogonal nuclear motions to alternative reactions.

Equivalent CASPT2 PE profiles of the ground and various excited states associated with the T + water reaction were also constructed. The returned profiles are analogous in topography to those of the U + water reaction and are therefore displayed in fig. S4 of the ESI. The energies associated with the FC geometry, as well as the energy of the equivalent S_1/S_0 CI are displayed as open points in fig. 6. As evident in fig. 6, the FC transitions are largely analogous in energy to that of U + water, whereas the CI is 0.3 eV less stable in T + water when compared to that of U + water. Such a destabilisation in the CI associated with the T + water reaction can be understood by considering that the initial PT reaction yields an electrophilic C6 position. The +I inductive effect associated with the methyl group at C5 serves to destabilise the biradical intermediate when compared to the analogous null inductive effect of H at the C5 position in the CI associated with U + water.

2.3 The thymine + water photoreaction in bulk DNA solution

In the forthcoming section we extend the above isolate gas-phase study to a somewhat complex biological environment. In so doing we use a model of double helix DNA solvated in bulk aqueous solution. The DNA model comprises strands of deoxyribose-adenine and thymine in which each strand comprises a common nucleobase - the equilibrium ensemble structure of which is depicted in fig. 7. Since we are interested in DNA, the mechanism will comprise T + H_2O in this section.

As with the gas phase, the state characters associated with the S_1, S_2 and S_3 are, respectively, nπ^*, $\pi\pi^*$ and $\pi\pi^*$. The returned states contain analogous orbital promotions to those of the gas-phase U/T + water complex.

The above studies on the T/U + water photoreaction show an initial path to a low energy CI - motion towards which predominantly involves H atom migration from water to C5-U/T. The ensuing reaction from the CI involves the addition of the nascent OH radical to C6-T/U - forming the photohydrate adduct. Given this two step reaction, two PE profiles were computed following ground state DFT/AMBER optimisations - which is outlined in more detail in the methodology. Briefly, the first comprises a relaxed scan in which the O-H stretch coordinate of water (henceforth R_{OH}) is varied, fixed at various values and the remainder of the nuclear framework allowed to relax (at the (TD)DFT/AMBER level - see methodology) to their respective minima on S_0 and S_1. This coordinate represents proton-transfer from the water molecule to T. The second step is modelled by varying the C6-OH stretch distance (henceforth R_{C6-OH}), and again fixing at various values and allowing the remaining nuclear degrees-of-freedom to relax (at the (TD)DFT/AMBER level - see methodology, ((TD)DFT denotes DFT or TDDFT)) to their respective minima. The re-

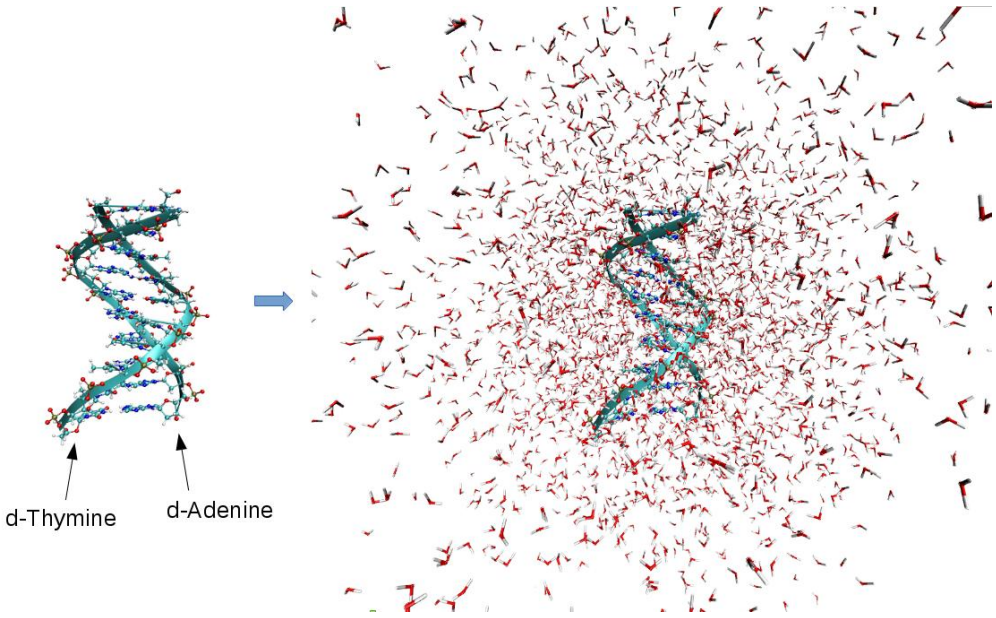


Fig. 7 Equilibrated starting structure of bulk DNA

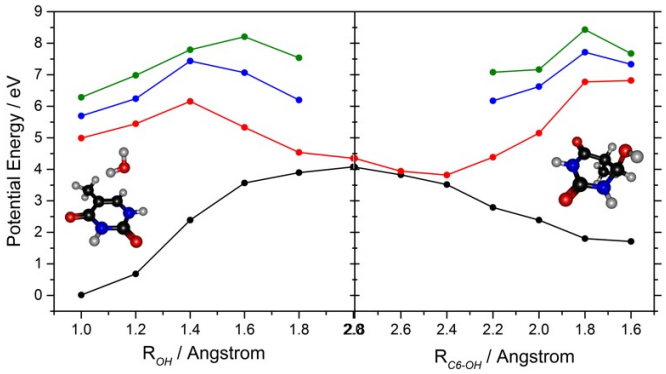


Fig. 8 CASPT2 PE profile for the T + water reaction in bulk solution. The S_0, S_1, S_2 and S_3 states are, respectively, the black, red, blue and green curves.

turned PE profiles associated with R_{OH} and R_{C6-OH} are presented in fig. 8. As shown, the profiles in the bulk biological environment bare certain resemblances (in topography and electronic state character) to those obtained in the gas phase (fig. 6 above) - reinforcing the ideology that the above description of gas-phase photoreaction is active in the bulk. As before, the ground state shows a sharp rise in PE as a function of R_{OH} - indicating a lack in driving force for PT from water to T. In the range $1.0 > R_{OH} > 1.3$ Å, the S_1, S_2 and S_3 states also show a rise in energy as a function of R_{OH} but become progressively reactive with respect to proton transfer at $R_{OH} > 1.3$ Å. This initial rise and then decline in PE, at respectively, $R_{OH} < 1.3$ Å and $R_{OH} > 1.3$ Å, is again a manifestation of the change in electronic state character from LE to CT. As before, the LE state contains orbital promotions that are intrinsic to the T chromophore whereas the CT states arises via an electron promotion from one of two O 2*P* orbitals localised on the water molecule to a π^* orbital local to T. The returned barrier top at $R_{OH} = 1.3$ Å (which again necessarily represents an upper limit) is 1 eV above that of the S_1 FC geometry. As with the aforementioned gas phase reaction, we note that S_1 is a dark $n\pi^*$ state, whilst the vertical S_2 state is a bright $\pi\pi^*$ state. The return maximum along the S_1 adiabatic path is approximately 0.2 eV above that of the vertical geometry of the bright S_2 state, showing a significantly smaller energy gap compared with that of the gas phase. Additionally, given that this inherent motion is an H atom motion, the long-range CT state may be populated by tunnelling from the LE state at near-threshold excitations. As before, the subsequent decline and rise in PE of, respectively, the S_1 and S_0 states inevitably manifests in a low-energy CI. From this low energy crossing point, the remainder of the PE profiles (right panel in fig. 8) shows a gradual decline in PE as a function of progressive shortening of C6-OH, whereas S_1 shows a net rise in PE. The ground state optimised structure of the photohydrate (right-most point in the right panel in fig. 8) is 1.9 eV less stable than the dispersively bound T + water complex. This is in contrast to the gas phase reactivity outlined in section 2.2, which shows that the relative stabilities of the photohydrate and the dispersively bound T/U + water complex are nigh on isoenergetic. This destabilisation in bulk solution may be a manifestion of the greater steric hinderance arising via the puckered ring geometry in the double-helix DNA environment.

Having constructed the reactivity of the T + water photoreaction in bulk DNA, future studies will compare and contrast the above described bulk T + H_2O photoreaction with those of the photocycloaddition reaction between adjacent T nucleobases. This is the sole reason for choosing DNA strands that contain common nucleobases on a single strand. In any future study we also expect to extrapolate the above QM/MM studies to U + H_2O.

3 General Discussion and Conclusions

Using high-level electronic structure calculations we have detailed the reaction between a given DNA/RNA nucleobase and

proximal water molecules - in an attempt at understanding the ways in which photo-induced nucleophilic hydrolysis reactions lead to harmful hydrate adducts. Within this class of photoreaction, the chosen nucleobases - G and T/U - represent well-known experimental cases in which G forms 8-OG and U/T form hydrate adducts. In the case of the G + water photoreaction we have shown that the well-known formation of 8-OG adducts can be formed via an initial photo-induced splitting of water - initiated via an EDPT from water to the N4 acceptor of G. The proton-transfer reaction leads to a protonated guanine + OH biradical which represents a low-energy crossing between the S_1 and S_0 states. IC at this crossing, followed by subsequent population of S_0, is predicted to give rise to a vibrationally-hot ground state biradical, with an energy in excess of 5 eV. We have shown that such an energy is in excess of the reaction barrier associated with the ultimate formation of 8-OG - which is predicted to proceed via an 8-HG intermediate formed via the addition of the nascent OH radical to the C2 position. As mentioned above, our theoretical studies complement the work of Ravanat and co-workers who have shown experimental evidence for the formation of 8-OG via a photoreaction of G with water.[13] Such a photoreaction also shows the capability of G to behave as a sensitiser for the initial step to water splitting in near- to mid- UV. In a prebiotic, ozone-free earth, an abundance of hard tropospheric UV radiation may have initiated the important first steps to photosynthesis via DNA/RNA nucleobase sensitisation. The well-known photorepair properties of 8-OG (coupled with the presently described labile photoreaction for forming 8-OG in the mid-UV) may have also been important in preserving the canonical DNA/RNA nucleobases in such a prebiotic environment.

Such photoreactions also extrapolate to U and T - both of which show a diverse range of differences and similarities to that of G. An important difference is the addition of H and OH across adjacent C-atoms in which the initial protonation occurs at C5. In the equivalent G + water reaction, the initial protoatation step is predicted to occur via proton-transfer to an N-acceptor local to G. Despite the differences in the nature of the acceptor atom, the potential energy landscape and the S_1/S_0 CI and state characters associated with the initial EDPT reaction are analogous to G + water. In all cases, the nascent protonated GH/UH/TH + OH biradical exhibits a 'downhill' ground-state energy profile as a function of the addition of OH to C6/N4. In the case of 8-HG, we have shown that H_2 elimination leads to the formation of 8-OG - the profile of which displays an energy barrier of approximately 3 eV. An equivalent H_2 elimination reaction is also expected following the formation of the UOH and TOH hydrate adducts. Such reactions have not been detailed experimentally thus are beyond the scope of this theoretical study.

In comparing the exited-state chemistry of U + water with T + water we note that the S_1/S_0 CI, associated with the TH + OH biradical, is less stable than that of the UH + OH biradical. This suggests a less stable CT state which likely manifests in the experimentally measured lower yield of thymine hydrate adducts when compared to uracil hydrate adducts.[15] Instead, it is well-known that photoinduced [2+2] cycloaddition reactions, between stacked adjacent thymine nucleobases, lead to the formation of thymine-thymine CPD lesion adducts. The equivalent photoreaction in stacked uracil nucleobases occurs with a significantly lower yield.[12] This observation may be reinforced by the presently obtained energetics which show a more favourable driving force for hydrate formation in U as compared to T. The expected larger quantum yield for uracil-hydrate adducts may quench the formation of U-U cycloadducts. In a fuller future study, we aim to compare and contrast the competition between cyclo- and hydrate- adduct formations in bulk DNA and RNA. Our present model DNA system is ideal for such a comparison since the thymine nucleobases are common to a single strand and are ideally positioned for a given [2+2]-cycloadddition study. In the present study, we use this bulk DNA model to compare the T + water reaction in a bulk environment to that of the gas phase. It should be noted that the quantum calculations in the gas phase and bulk environment are not equivalent (as will be discussed in the Methods section the active space in bulk is smaller for computational reasons), so an exact quantitative comparison is not appropriate. We can make qualitative comparisons however. In comparing the gas and condensed phase reactivity of the T + water reaction we note that the PE profiles between the two phases are qualitatively similar. Both show an adiabatic S_1 PE profile that contains a net reactivity with respect to proton-transfer whilst that of S_0 shows no net reactivity. The S_1/S_0 CI at long range R_{0H} is also structurally similar in both the gas phase and in bulk solution. In contrast, two key energetic differences are noted. The first is the relative stability of the hydrate-adduct, which in the bulk environment, is predicted to be approximately 1.9 eV less stable than in the gas phase. The second is the noteworthy energy difference, between the bright S_2 state and the local maximum on the S_1 adiabatic profile, which is significantly smaller in the bulk DNA environment compared with that of the gas phase. In bulk solution, this smaller energy difference leads to an enhancement in the competition between the photohydrate forming reaction and the non-radiative (nucleobase preserving) decay paths that are intrinsic to the T nucleobase.

With the above discussion in mind, detailed studies of this class of water-chromophore reaction in the gas phase - in which the unperturbed chromophore-complex intrinsic excited-state reactivity is qualified - is an important stepping stone for extrapolation into the bulk. In so doing, our present study shows that despite the minor energetic differences, the excited-state reaction is equally as active in a complex bulk environment as is in the gas phase leaving the door open for possible experiments of nucleobases water clusters in the gas phase. Such experiments are already showing promise on related chromophore-water clusters[27]. In conclusion we have shown that such hydrate adducts can compete with other well-known lesion-forming adducts and more importantly, with the photostabilising non-radiative decay paths.

4 Methods

4.1 The Guanine + Water reaction

The 9H-guanine molecule was clustered in an ensemble of five solvent water molecules. The five cluster molecules of water were selected in order to optimally describe the significant dis-

persive and hydrogen-bonding effects required for the ultimate formation of a hydrate adduct in which the OH adds to C2. Such a cluster is also motivated from previous studies on analogous systems.[23,24,28] Five water molecules also represent the limit at which we can obtain a detailed qualitative description of the excited-state energetics, intrinsic to this specific class of water-chromophore reaction, whilst maintaining a decent computational expense. Relaxed geometries, as a function of the R_{OH} driving coordinate of a H2O-N7 centred hydrogen-bond, were obtained by fixing R_{OH} at various values and optimising the remaining internal degrees-of-freedom using RI-ADC(2)[29]/cc-pVDZ[30] and RI-MP2[31]/cc-pVDZ[30] for, respectively, the S_1 and S_0 states. The final energies associated with returned relaxed MP2/ADC(2) profiles were computed using the complete active space with second-order perturbation theory[32,33] (CASPT2) based on a state-averaged complete acitve space self-consistent field reference wavefunction[34] (SA4-CASSCF). The active space was carefully chosen to strike a balance between an adequate description of the excited-state chemistry and a reasonable computational expense. After careful test, an optimal active space of ten electrons in ten orbitals (3 π, 3 π^*, 2 O(2*P* lone-pair, 2σ^* orbitals) was used. An imaginary level shift of 0.5 E_H was used to aid convergence and to mitigate the involvement of intruder states. All ADC(2) and MP2 calculations were undertaken in Turbomole v2.2[35] whilst all CASPT2/CASSCF calculations were undertaken in Molpro 2015.1[36].

4.2 The uracil/thymine + water reaction in gas phase

The ground state minimum energy geometry of the uracil-water complex, thymine-water complex and the uracil and thymine hydrate adducts were optimised using the Becke 3-parameter Lee-Yang-Parr functional[37–40] of Density Functional Theory, coupled to the 6-31G(d) Pople basis set[41] (DFT/B3LYP/6-31G(d)). The subsequent real 3N-6 normal-mode wavenumbers confirmed that all optimised structures were at true minima. Transition states, connecting the uracil/thymine + water complex to the hydrate adducts, were optimised using the Synchronous Transit-Guided Quasi-Newton method[42] at the DFT/B3LYP/6-31G(d) level of theory. The returned transition state was used to compute an intrinsic reaction path (IRC) in order to obtain the minimum energy path connecting the reactions to the products. CIs between the S_0 and S_1 state were located for both uracil/water and thymine/water, using the SA3-CASSCF method, coupled to a 6-31G(d) Pople basis set. The CIs so derived were computed using an active space of six electrons in six orbitals. The PE profile connecting the ground state minimum energy geometries to the minimum energy CIs were obtained by means of LIIC and thus necessarily represent an upper bound to the true energetic.The energies along the LIIC were obtained with CASPT2, based on an SA6-CASSCF reference wavefunction and a full-valence active space of twenty electrons in fourteen orbitals. As with the guanine + water reaction, an imaginary level shift of 0.5 E_H was used in all CASPT2 calculations. Energies along the LIIC were also computed using Time-Dependent Density Functional Theory[43–45] (TDDFT/B3LYP/6-31G(d) in order to qualitatively compare the topography of the PE profile to that returned by CASPT2/CASSCF(20,14)/6-31G(d). The reactant and product excitation energies were also calculated using the Equation-of-motion Coupled Cluster Singles and Double method[46,47] (EOM-EE-CCSD), coupled to a 6-311G(d) Pople basis set. The (TD)DFT and EOM-EE-CCSD calculations were performed in Q-Chem[48]. The CIs were located using Columbus 7.0[49–53] and the CASPT2/CASSCF(20,14) single point calculations were performed using Molpro 2015.1[36]. The transition state optimisations and the subsequent IRC were obtained using Gaussian 09[54].

4.3 QM/MM methods for the T + water photoreaction

Using AmberTools v16[55], a double-helix structure of B-DNA, comprising ten deoxyriboseadenine and deoxyribosethymine (in which each strand contained a common nucleoside) was constructed and solvated with a TIP3P ensemble of solvent water molecules in a truncated octahedron box. Na^+ counter ions were used to neutralise the net negative charge provided by the phosphate moieties. The AMBER ff99 parameter set[56] (in combination with ff99bsc0 corrections[57] was adopted in the present study and has been shown to work well in analogous systems.[58,59] Using the Sander module in Amber 16,[60] the subsequent structure was minimised for 2,500 steps in order to mitigate any steric clashes and then gradually annealed from 0 K to 298 K for 20,000 steps with a time step of 1 fs. Positional restraints with the force constant of 25 kcal mol^{-1} Ang^{-2} were applied on the solute during heating. At 298 K, the applied restraint was gradually removed in 5 increments. A Langevin thermostat, with a collisional frequency of 5 ps^{-1}, was used in order to maintain a constant temperature. A constant pressure of 1 bar was maintained by applying a weak-coupling algorithm with a relaxation time of 1 ps. A particle mesh Ewald summation method was used for calculating long-range electrostatic interactions. The SHAKE algorithm was used to constrain all degrees-of-freedom associated with hydrogen-atom motions. The system was equilibrated for 1 ns prior to the QM/MM calculations.

The mixed quantum and classical (QM/MM) calculations were performed using our in-house QutELa (**Qu**an**t**um **E**lectronic structure with **La**yered molecular mechanics) programme in which Gaussian or Molpro (for the quantum QM region) and Amber 16 (for the classical MM region) were coupled in order to obtain QM/MM minima and associated energetics. The QM region comprised a single thymine molecule at the end of the strand and the 'nearest-neighbour' water molecule returned from the equilibration. The remaining residues and water molecules made up the classical MM region. A standard additive QM/MM Hamiltonian, as displayed in equation 1, is solved in order to determine the total energy of the bulk system. The terms in equation 1 represent the isolated QM part (H_{QM}) describing the QM atoms, an isolated MM part (H_{MM}) describing the MM atoms and a coupling term ($H_{QM/MM}$) describing the intrinsic coupling between the QM and MM regions. In the present case $H_{QM/MM}$ is introduced by augmenting the zero-order electronic Hamiltonian H_{QM} of the QM region with an additional potential energy term as given in equa-

tion 2. The first term in equation 2 represents the electrostatic effect that an atom M with charge Q_M at R_M (in the MM region) has on the electronic coordinates r_i of the solute. The second term represents the electrostatic interaction between an atom N with charge Z_N at R_N (in the QM region) with a point charge Q_M at R_M. Polarisability and Van der Waals terms are ignored in $H_{QM/MM}$ as the former has been shown to introduce minor effects upon the returned energetics in related systems whilst the latter is independent of the electronic coordinates of the solute represented by the QM region.[61,62] A link hydrogen atom, connecting the QM and MM regions, was introduced in the QM region in order to maintain valency. The point charge associated with the MM atom associated with the link-atom was set to zero in order to prevent over-polarisation.

$$H_{Total} = H_{QM} + H_{MM} + H_{QM/MM} \quad (1)$$

$$H_{QM/MM} = \sum_{iM} \frac{Q_M}{r_i - R_M} - \sum_{MN} \frac{Z_N Q_M}{R_M - R_N} \quad (2)$$

For the initial proton transfer step, relaxed geometries along the S_0 and S_1 states were obtained by using the R_{OH} coordinate of the proximal water molecule (of the thymine + water complex) as the driving coordinate; in which R_{OH} was fixed at various values and the remainder of the nuclear framework associated with the QM and MM regions were allowed to relax at the (TD)DFT/ωB97XD/6-31G(d):AMBER level of theory (coupling Gaussian 09 with Amber v16.0). This relaxation was performed by using the well-known micro-optimisation method for the MM region. The ωB97XD[63] functional was chosen since the present system requires the optimal description of short- and long- range correlation as well as dispersion interactions. This same procedure was repeated for the second reaction step in which the nascent OH radical adds to C6 but this time by using R_{C-OH} as the driving coordinate. Single-point energies associated returned relaxed geometries along R_{OH} and R_{C-OH} were then recomputed using CASPT2. The CASPT2 calculations were performed based on a SA4-CASSCF reference wavefunction and comprised an active space of ten electrons in ten orbitals (3 π, 3 π^*, 2 σ^* and 2 O($2P_Y$)). An imaginary level shift of 0.5 E_H was used.

5 Acknowledgements

The authors thank Carina Tse for driving the earlier studies on the U/T + water photoreaction. P.C., T.N.V.K. and S.M. gratefully acknowledge the National Science Foundation (NSF, grant number: CHE-1465138) for funding. B.M. gratefully acknowledges the European Union Horizon 2020 research and innovation programme for the award of a Marie Sklodowska Curie fellowship (PhARRAO initiative; grant number: 746593).

References

1 C. E. Crespo-Hernandez, B. Cohen, P. M. Hare and B. Kohler, *Chem. Rev.*, 2004, **104**, 1977–2019.

2 C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernandez and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217–239.

3 R. Improta, F. Santoro and L. Blancafort, *Chem. Rev.*, 2016, **116**, 3540–3593.

4 *Topics in Current Chemistry - Photoinduced Phenomena in Nucleic Acids I*, ed. S. Ullrich, A. C. Borin and M. Barbatti, Springer, Berlin - Heidelberg, 2015, vol. 355.

5 *Topics in Current Chemistry - Photoinduced Phenomena in Nucleic Acids II*, ed. S. Ullrich, A. C. Borin and M. Barbatti, Springer, Berlin - Heidelberg, 2015, vol. 356.

6 T. Gustavsson, R. Improta and D. Markovitsi, *J. Phys. Chem. Lett.*, 2010, **1**, 2025–2030.

7 S. Mai, M. Richter, P. Marquetand and L. González, *Top. Curr. Chem.*, 2015, **356**, 99–153.

8 W. Domcke, D. R. Yarkony and H. Köppel, *Conical Intersections*, World Scientific, Singapore, 2004.

9 S. Matsika and P. Krause, *Annu. Rev. Phys. Chem.*, 2011, **62**, 621–643.

10 A. L. Sobolewski and W. Domcke, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4897.

11 *CRC Handbook of Organic Photochemistry and Photobiology*, CRC Press, New York, 1995.

12 C. L. Greenstock, I. H. Brown, J. W. Hunt and H. E. Johns, *Biochem. Biophys. Res. Commun.*, 1967, **27**, 431–436.

13 M. Gomez-Mendoza, A. Banyasz, T. Douki, D. Markovitsi and J.-L. Ravanat, *J. Phys. Chem. Lett.*, 2016, **7**, 3945–3948.

14 J. G. Burr, E. H. Park and A. Chan, *J. Am. Chem. Soc.*, 1972, **94**, 5866.

15 H. Görner, *J. Photochem. Photobiol. B: Biol.*, 1991, **10**, 91–110.

16 J. Remsen, M. Mattern, N. Miller and P. Cerutti, *Biochem.*, 1971, **10**, 524–529.

17 T. Ganguly and N. J. Duker, *Nucleic Acids Res.*, 1991, **19**, 3319–3323.

18 R. J. Boorstein, T. P. Hilbert, J. Cadet, R. P. Cunningham and G. W. Teebor, *Biochem.*, 1989, **28**, 6164–6170.

19 S. Franzen, B. Skalski, L. Bartolotti and B. Delley, *Phys. Chem. Chem. Phys.*, 2014, **16**, 20164–20174.

20 T. Ganguly and N. J. Duker, *Mutat. Res./DNA Repair*, 1992, **293**, 71–77.

21 R. Szabla, R. W. Gora, M. Janicki and J. Sponer, *Faraday Discuss.*, 2016, **195**, 237–251.

22 R. Szabla, D. Tuna, R. W. Góra, A. L. S. J. Sponer and and W. Domcke, *J. Phys. Chem. Lett.*, 2013, **4**, 2785–2788.

23 X. Wu, T. N. V. Karsili and W. Domcke, *ChemPhysChem*, 2016, **17**, 1298–1304.

24 M. Barbatti, *J. Am. Chem. Soc.*, 2014, **136**, 10246–10249.

25 G. Fisher and H. Johns, in *Pyrimidine Photohydrates*, ed. S.-Y. Wang, 1976, vol. 1, pp. 169–224.

26 A. Yoshikawa and S. Matsika, *Chem. Phys.*, 2008, **347**, 393–404.

27 N. Esteves-Lopez, S. Coussan, C. Dedonder-Lardeux and C. Jouvet, *Phys. Chem. Chem. Phys.*, 2016, **18**, 25637–25644.

28 R. Szabla, J. Sponer and R. W. Góra, *J. Phys. Chem. Lett.*, 2015, **6**, 1467–1471.

29 A. B. Trofimov and J. Schirmer, *J. Phys. B: At. Mol. and Opt. Phys.*, 1995, **28**, 2299–2324.
30 T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007–1023.
31 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359 – 363.
32 B. O. Roos and K. Andersson, *Chem. Phys. Lett.*, 1995, **245**, 215 – 223.
33 K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski, *J. Phys. Chem.*, 1990, **94**, 5483–5488.
34 B. O. Roos, P. R. Taylor and P. E. Sigbahn, *Chem. Phys.*, 1980, **48**, 157 – 173.
35 R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.
36 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and M. Wang, *MOLPRO, version 2015.1, a package of ab initio programs*, 2015, see.
37 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
38 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
39 S. H. Vosko, L. Wilk and M. Nusair, *Canadian J. Phys.*, 1980, **58**, 1200–1211.
40 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
41 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.
42 C. Peng and H. B. Schlegel, *Israel J. Chem.*, 1993, **33**, 449–454.
43 E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.
44 R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.*, 1996, **256**, 454 – 464.
45 F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2002, **117**, 7433–7447.
46 J. F. Stanton and R. J. Bartlett, *J. Chem. Phys.*, 1993, **98**, 7029–7039.
47 A. I. Krylov, *Ann. Rev. Phys. Chem.*, 2008, **59**, 433–462.
48 Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kus, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stueck, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184–215.
49 H. Lischka, T. Müller, P. G. Szalay, I. Shavitt, R. M. Pitzer and R. Shepard, *Wiley Interdisciplinary Reviews: Comput. Mol. Sc.*, 2011, **1**, 191–199.
50 H. Lischka, R. Shepard, R. M. Pitzer, I. Shavitt, M. Dallos, T. Muller, P. G. Szalay, M. Seth, G. S. Kedziora, S. Yabushita and Z. Zhang, *Phys. Chem. Chem. Phys.*, 2001, **3**, 664–673.
51 H. Lischka, R. Shepard, I. Shavitt, R. M. Pitzer, M. Dallos, T. Müller, P. G. Szalay, F. B. Brown, R. Ahlrichs, H. J. Böhm, A. Chang, D. C. Comeau, R. Gdanitz, H. Dachsel, C. Ehrhardt, M. Ernzerhof, P. Höchtl, S. Irle, G. Kedziora, T. Kovar, V. Parasuk, M. J. M. Pepper, P. Scharf, H. Schiffer, M. Schindler, M. Schüler, M. Seth, E. A. Stahlberg, J.-G. Zhao, S. Yabushita, Z. Zhang, M. Barbatti, S. Matsika, M. Schuurmann, D. R. Yarkony, S. R. Brozell, E. V. Beck, J.-P. Blaudeau, M. Ruckenbauer, B. Sellner, F. Plasser and J. J. Szymczak, *COLUMBUS, an ab initio electronic structure program, release 7.0 (2012)*.
52 H. Lischka, M. Dallos, P. G. Szalay, D. R. Yarkony and R. Shepard, *J. Chem. Phys.*, 2004, **120**, 7322–7329.
53 M. Dallos, H. Lischka, R. Shepard, D. R. Yarkony and P. G. Szalay, *J. Chem. Phys.*, 2004, **120**, 7330–7339.
54 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi,

N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision B.01*, 2009.

55 D. Case, D. Cerutti, T. Cheatham, III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, C. Simmerling, W. Botello-Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, L. Xiao, D. York and P. Kollman, *Amber 17*, 2017.

56 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.

57 A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817 – 3829.

58 W. Lee and S. Matsika, *Phys. Chem. Chem. Phys.*, 2017, **19**, 3325–3336.

59 W. Lee and S. Matsika, *Phys. Chem. Chem. Phys.*, 2015, **17**, 9927–9935.

60 D. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Götz, I. Kolossváry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. Kollman, *AMBER 12, University of California, San Francisco*, 2012.

61 Z. Xu and S. Matsika, *J. Phys. Chem. A*, 2006, **110**, 12035–12043.

62 J. Kongsted, A. Osted, K. V. Mikkelsen, P.-O. Åstrand and O. Christiansen, *J. Chem. Phys.*, 2004, **121**, 8435–8445.

63 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.